(* COUNTS *)

29y

Less simple examples

		Exh.	Page
Introduction 29y-1			
		A	
		1	29y-2
29yA.	Ages of married couples 29y-1	B	
		2	29y-6
29yB.	Contouring 29y-3	3	29y-7
	- cross-sections 29y-5	4	29y-9
	- intersections -- and level curves 29y-11	5	29y-10
	- going on 29y-13	6	29y-12
		7	29y-13
29yC.	"Smooth curving" a fit 29y-13	C	
	- end values 29y-20	8	29y-14/15
		9	29y-17/18
29yD.	Application to married couples 29y-24	10	29y-21/22
		11	29y-23
29yE.	Another look at the school girls 29y-37	D	
		12	29y-26/27
	**********	13	29y-28
		14	29y-29
29yH.	How far have we come? 29y-47	15	29y-30
		16	29y-31
	**********	17	29y-32
		18	29y-34
29yP.	Problems for chapter 29y 29y-49	19	29y-35
		20	29y-36
		E	
		21	29y-38
		22	29y-39/40
		23	29y-41/42
		24	29y-44/45
		H	
		P	

29y

Less simple examples

This short chapter supplements chapter 29x with some further examples that are not quite as simple as those discussed there.

29yA. Ages of married couples

The British census gives in 2-way tables the ages of husbands and wives enumerated on the same schedule -- those dwelling in the same house or apartment. While it seems to be stretching a point to take such pairs of ages as representing paired responses, what was recorded by the census takers are responses in a reasonable sense and are surely ordered. We can try the techniques of chapter 29x and see what they show us. As an example of not quite simple behavior, such an analysis is quite illuminating. Panel A of exhibit 1 summarizes the basic counts. The first column, for example, shows an unusual appearance -- first falling off rapidly and then slowing down, so that there are more wives in their 60's and 70's than one would suspect -- for husband's under 25. As Ansley Coale told us, this phenomenon is well-known to students of such data and is reasonable certainly due to classifying mothers as wives, etc. When we turn to the table of double flogs, we see a clear tendency for a valley almost, but not quite, down the diagonal,

exhibit 1 of chapter 29y -- married couples

Ages of husbands and wives enumerated on the same schedule

A) The COUNTS

age of wife		age of husband: to 24	25-29	30-39	40-49	50-59	60-69	70-79	80 up
			b	c	d	e	f	g	h
to 19	a	35118	12581	2320	220	50	4	6	2
20-29	c	265791	808564	660415	43891	4125	784	113	55
30-39	d	4922	115460	1595508	735016	50391	7167	1298	173
40-49	e	349	2991	162648	1694546	547684	54653	9096	851
50-59	f	63	286	5133	160930	1238366	406926	44052	3844
60-69	g	20	70	504	5596	120987	754186	208186	11345
70-79	h	10	12	75	455	3764	69744	295858	38801
80 up		11	11	65	54	268	1598	15012	27159

B) The DOUBLE (and SINGLE) FLOGS -- unit = 0.01

	b	c	d	e	f	g	h	(flog)
a	194	213	250	270	291	276	231	(-231)
c	248	203	210	250	279	295	251	(-65)
d	299	280	214	210	235	252	243	(-13)
e	319	316	288	223	205	214	219	(30)
f	314	330	318	278	222	196	197	(75)
g	284	319	317	309	271	223	201	(135)
h	209	246	248	272	267	252	247	(236)
(flog)	(-151)	(-85)	(-25)	(20)	(63)	(117)	(209)	((10260272))

Median: 250. Hinges: 216h and 282. 8ths: 207 and 315. Outs: 0/49

exhibit 1 (cont'd.)

C) Panel A in CODED FORM

	b	c	d	e	f	g	h
a	0	0	.	.	x	.	.
b	.	0	0	.	.	x	.
d	x	.	0	0	.	.	.
e	x	x	x	.	0	0	0
f	x	x	x	x	.	0	0
g	x	x	x	x	.	.	0
h	.	.	.	.	.	.	.

where

column flog = row flogs

paralleled on either side by a ridge about a generation's length -- say 30 years -- away.

These phenomena -- somewhat more clearly shown in coded form in panel C -- must correspond to the same sort of misclassification.

29yB. Contouring

With ten million pairs of ages in our table, irregularities due to the limited size of sample have relatively minor consequences. If we had to deal with a 1-way situation, we would expect to be able to plot points and draw a smooth curve through them -- rather than near them. What can we do in a 2-way situation that would correspond to this?

If we have a smooth surface to visualize, we have two choices:

- we can make a solid model -- using clay, wax, or some other material. After very considerable labor we will be able to look at the result from many directions. We can also handle and feel it.

- if we are to use paper, and -- we hope -- considerably less effort we can make a contour map.

With practice, especially if we practice on contour maps for areas of landscape and other surfaces familiar to us, we can do a quite reasonable job of appreciating a surface from a contour map. One easy way to approach an appreciation of a contour map is to think of various stages of flooding of our surface by standing water. Sea coast dwellers can think of various stages of the tide. Those more familiar with ponds and backwaters from streams can think first about slower changes in water level. Desert dwellers may want to think of pouring water into a vessel containing a model of the surface. In each case, there will be a shore line -- a set of pieces of a level curve -- for each lequid level. If we know all the level curves -- both as to where they go and at what elevation they lie -- we know the whole surface. In practice, we can see what goes on more simply from a selected set of level curves spaced neither too closely nor too widely.

In practice we almost always space such curves equally in the vertical. If we need to see small differences in level in one part of a surface without confusing very large differences elsewhere, however, we are of course free to use different steps at different heights -- or even in different regions -- so long as we give due warning.

What techniques do we need to use to pass from a two-way table of heights to a set of level curves? A good answer

to this question depends on -- at least -- the nature of the data and the means we use to go from data to contour map. If the data fills all -- or a reasonably connected part -- of a column-and-row table, and if we are to work by hand, the technique of drawing smooth cross-sections and marking points on appropriate level curves is the natural place to begin.

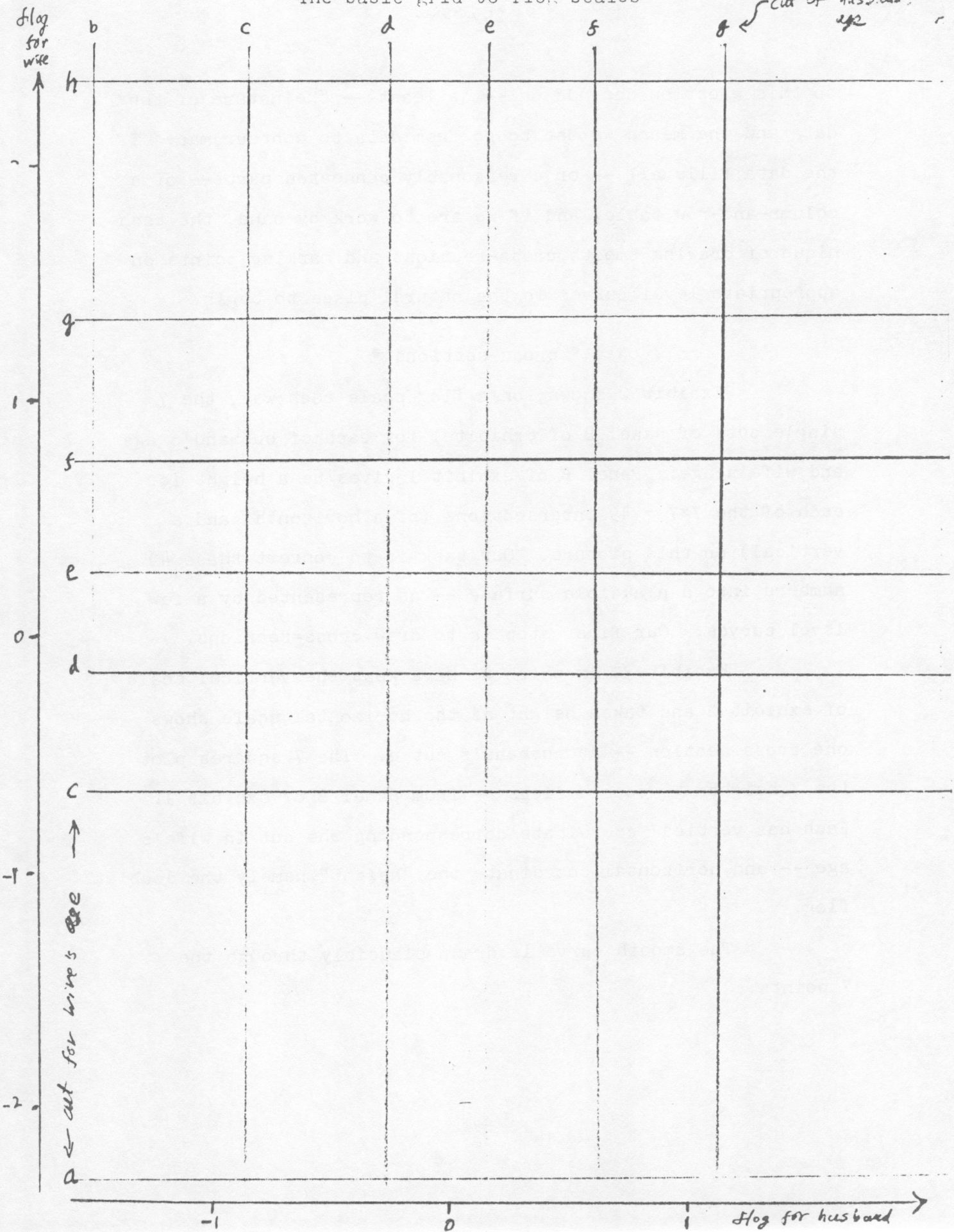
* cross-sections *

Exhibit 2 shows, on a flog scale each way, the 7 single cuts of panel B of exhibit 1 for each of husband's age and wife's age. Panel B of exhibit 1 gives us a height for each of the 7×7 = 49 intersections (of a horizontal and a vertical) on this picture. Our task is to convert these 49 numbers into a plausible surface -- as represented by a few level curves. Our first step is to draw cross=sections.

Exhibit 3, in which we have kept the vertical scale of exhibit 2 and taken height as the horizontal scale shows one cross section -- at husband's cut e. The 7 squares plot the 7 heights -- double flogs -- from panel B of exhibit 1. Each has vertical coordinate corresponding the cut in wife's age -- and horizontal coordinate the "height" namely the double flog.

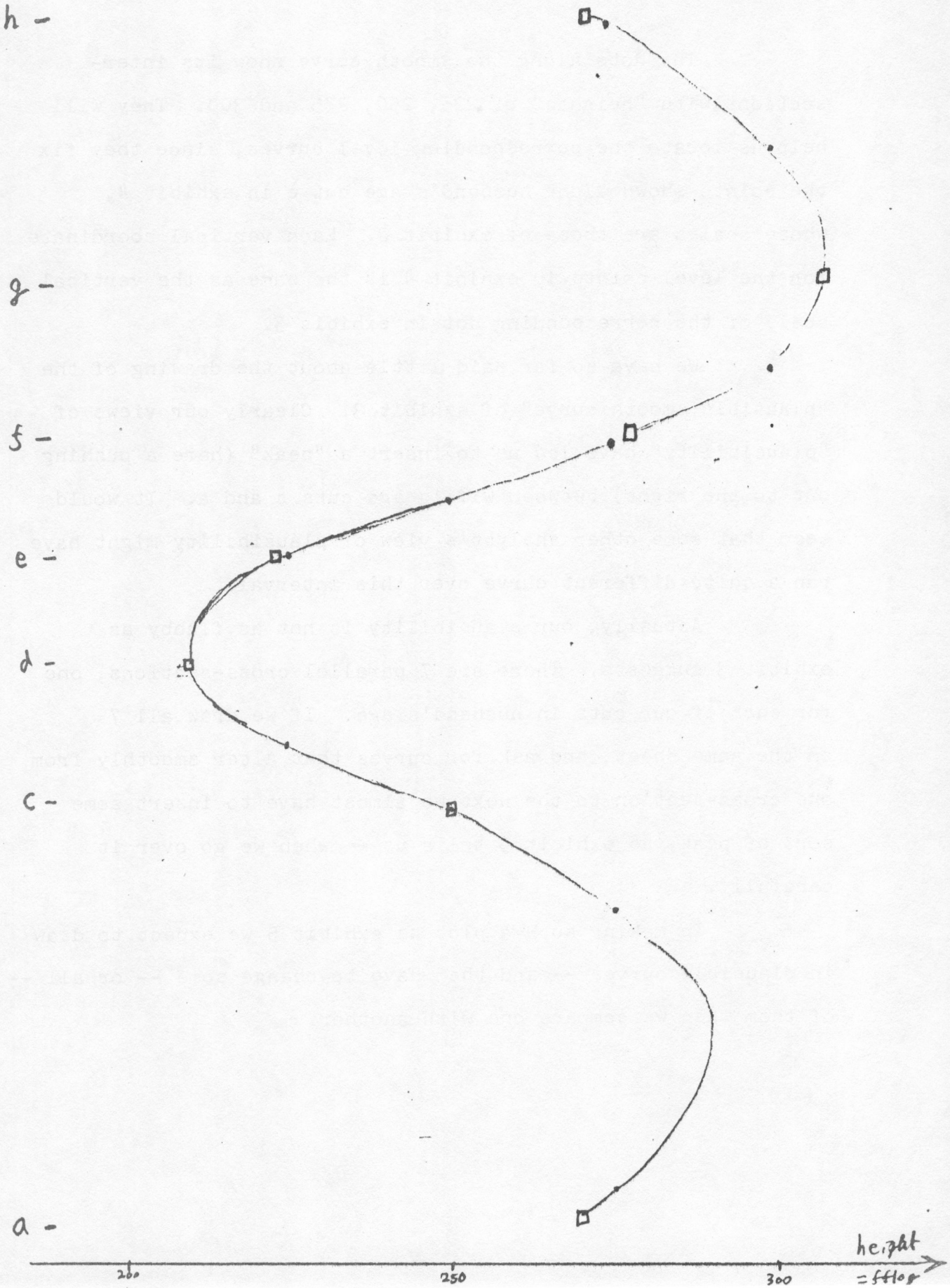
The smooth curve is drawn plausibly through the 7 points.

The basic grid to flog scales



One cross section: husband's age cut e
□ = data from panel B of exhibit 1
curve drawn smoothly through these 7 points

Cut-in wife's age



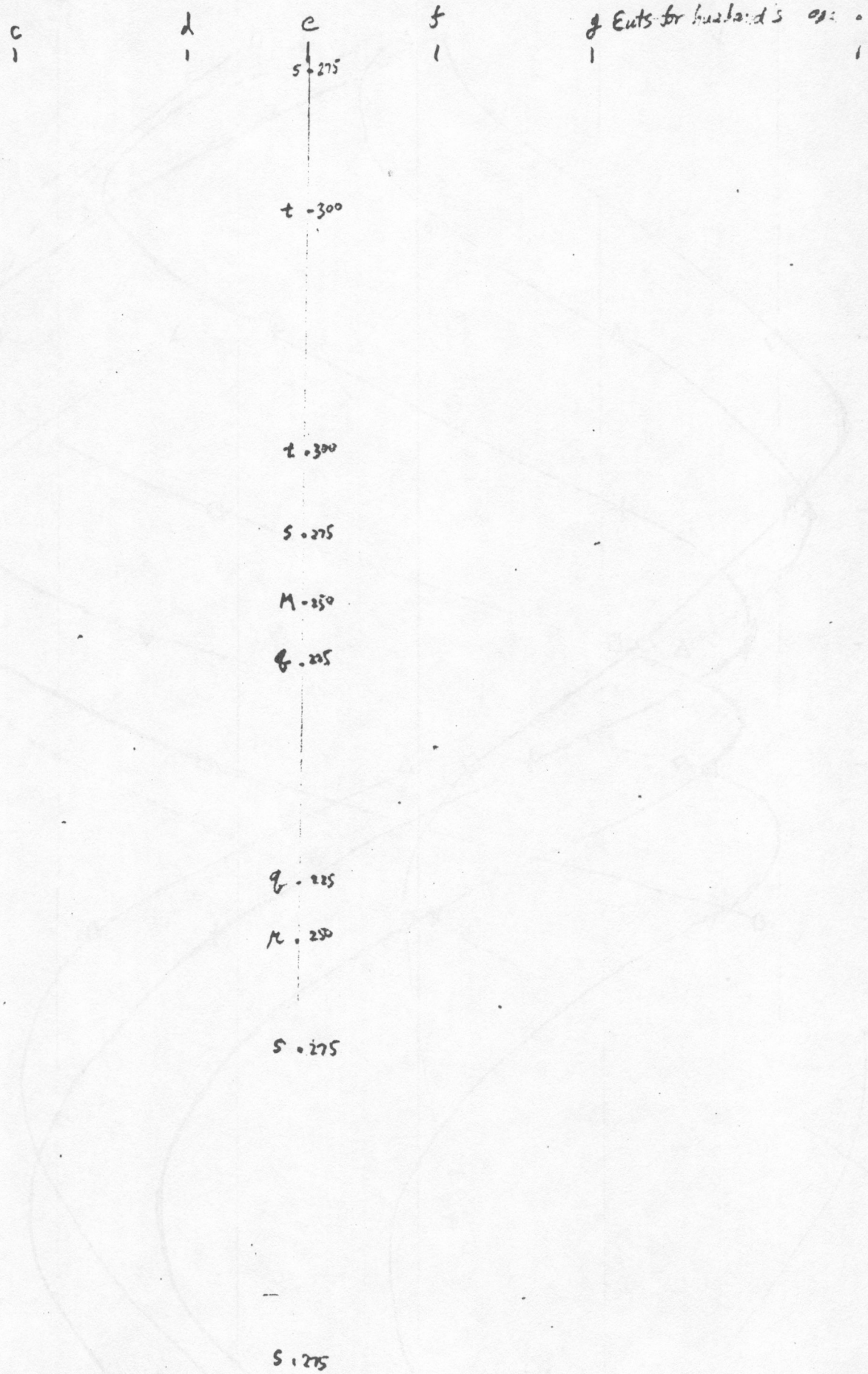
The dots along the smooth curve show its intersections with "heights" of 225, 250, 275 and 300. They will help us locate the corresponding level curves, since they fix the points shown along husband's age cut e in exhibit 4, whose scales are those of exhibit 2. Each vertical coordinate for the level points in exhibit 4 is the same as the vertical scale of the corresponding dot in exhibit 3.

We have so far said little about the drawing of the "plausible smooth curve" of exhibit 3. Clearly our views of "plausibility" have led us to insert a "peak" (here a pushing out to the right) between wife's age cuts c and a. It would seem that some other analyst's view of plausibility might have run a quite different curve over this interval.

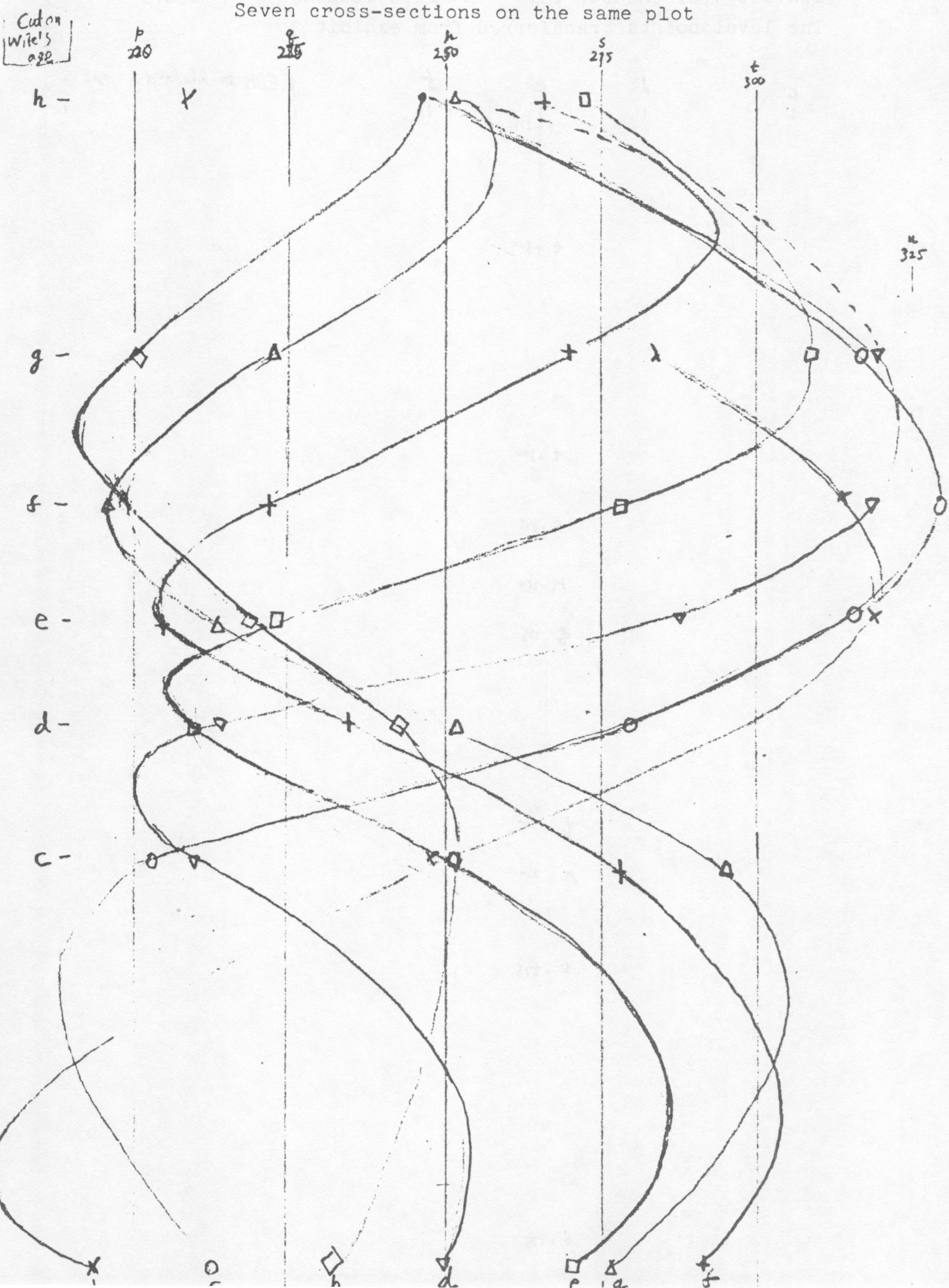
Actually, our plausibility is not as flabby as exhibit 3 suggests. There are 7 parallel cross-sections, one for each of our cuts in husband's age. If we draw all 7 on the same sheet, and ask for curves that alter smoothly from one cross-section to the next we almost have to insert some sort of peak, as exhibit 5 tells us -- when we go over it carefully.

In making such a plot as exhibit 5 we expect to draw in plausible curves -- and then have to change some -- or all -- of them when we compare one with another.

The level points transferred from exhibit 3



Seven cross-sections on the same plot



In exhibit 5 we have drawn in lightly the vertical lines that correspond to the level curves that we propose to construct. Each is labelled with its "height" in numbers and with a one-letter tag.

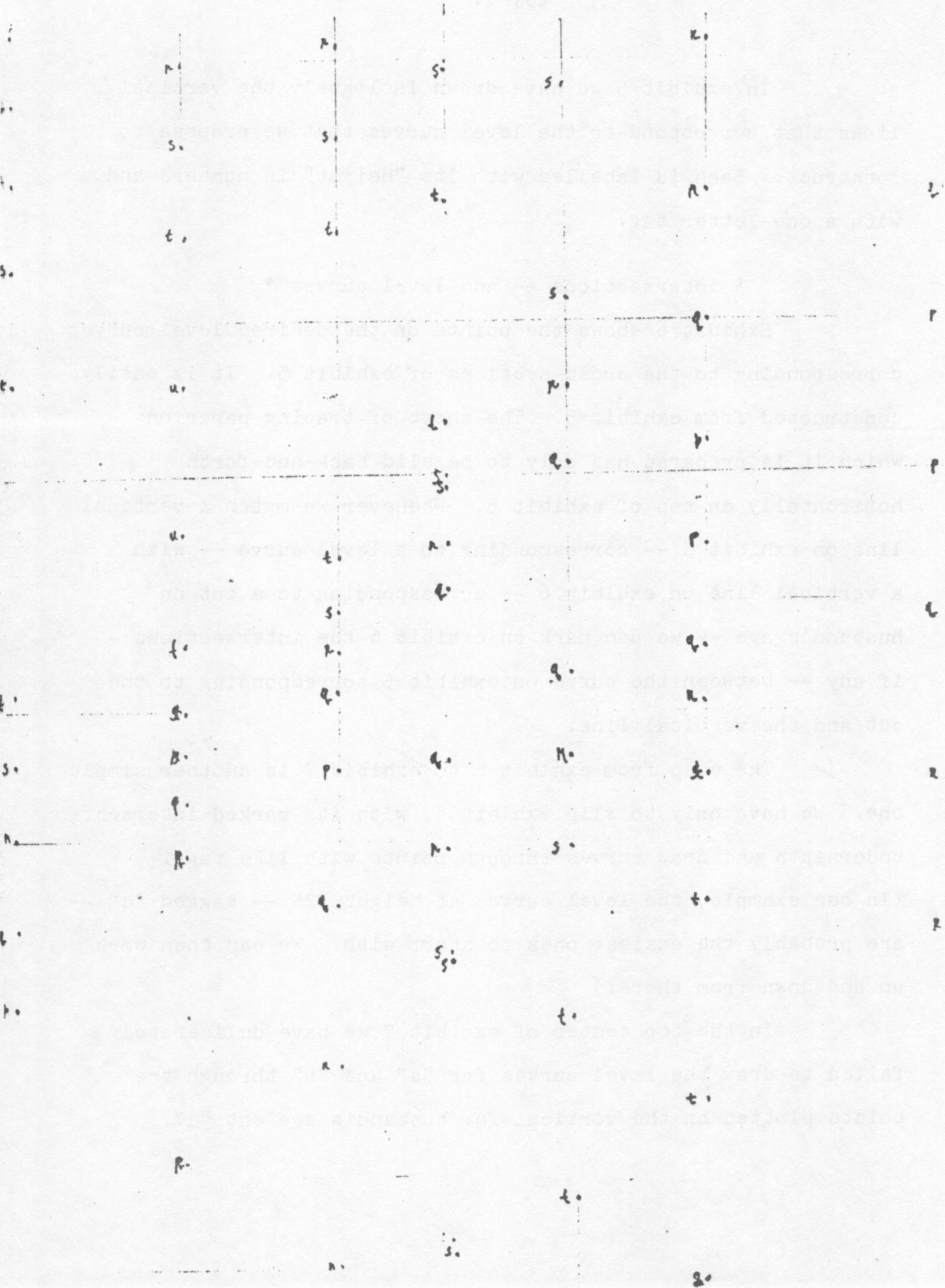
* intersections -- and level curves *

Exhibit 6 shows the points on the desired level curves corresponding to the cross-sections of exhibit 5. It is easily constructed from exhibit 5. The sheet of tracing paper on which it is prepared has only to be slid back-and-forth horizontally on top of exhibit 5. Whenever we match a vertical line on exhibit 5 -- corresponding to a level curve -- with a vertical line on exhibit 6 -- corresponding to a cut on husband's age -- we can mark on exhibit 6 the intersections -- if any -- between the curve on exhibit 5 corresponding to the cut and the vertical line.

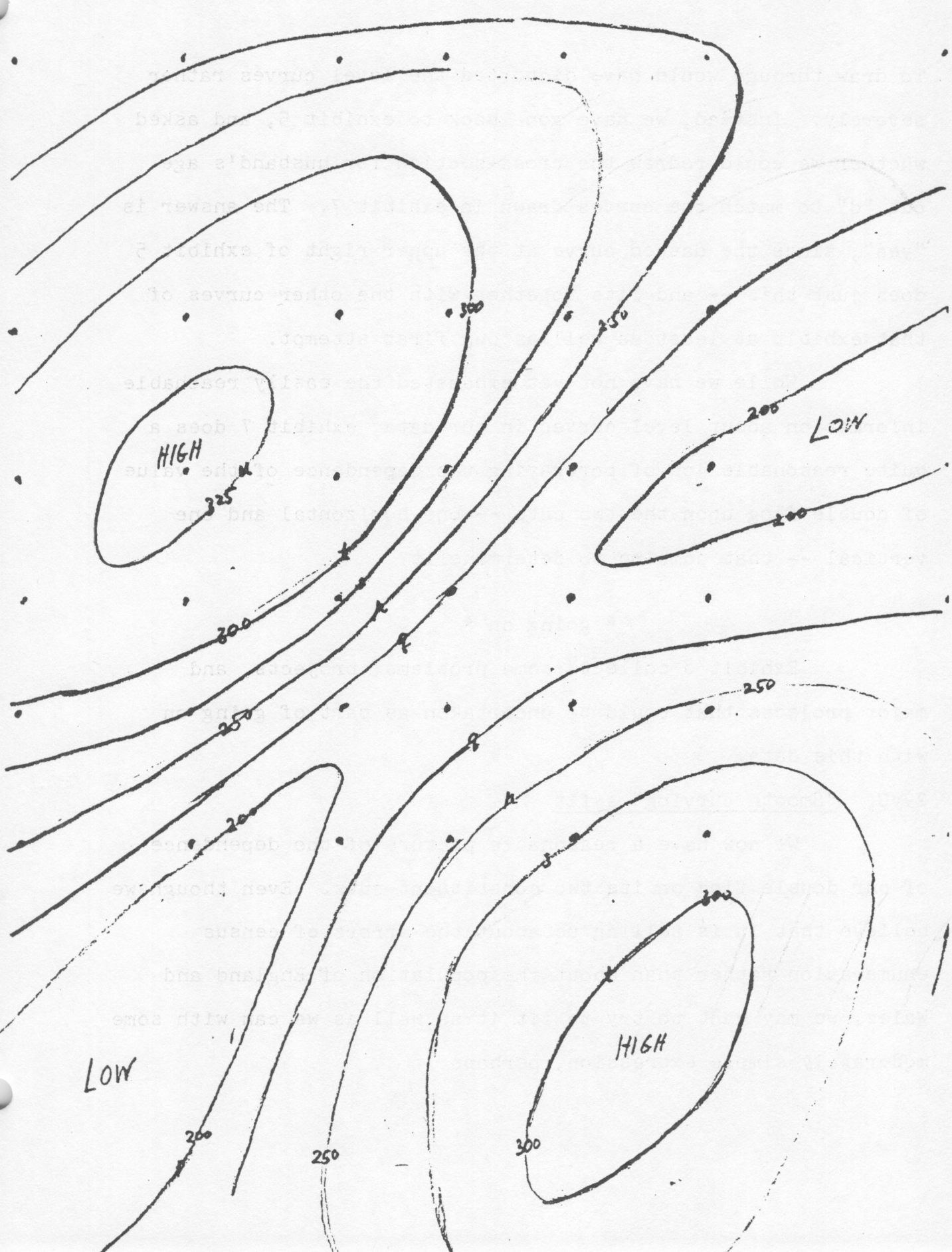
The step from exhibit 6 to exhibit 7 is another simple one. We have only to slip exhibit 6, with its marked intersections underneath and draw curves through points with like tags. (In our example, the level curves at height 225 -- tagged "q" -- are probably the easiest ones to start with. We can then work up and down from there.)

In the top center of exhibit 7 we have deliberately failed to draw the level curves for "s" and "t" through the points plotted on the vertical for husband's age cut "d".

Intersections of selected level curves with verticals



Contours (level curves) based on exhibit 8
(y-by-7 array of dots show sites of data)



To draw through would have distorted the level curves rather severely. Instead, we have gone back to exhibit 5, and asked whether we could redraw the cross-section for husband's age cut "d" to match the curves drawn in exhibit 7. The answer is "yes", since the dashed curve at the upper right of exhibit 5 does just this -- and fits together with the other curves of that exhibit at least as well as our first attempt.

While we have not yet exhausted the easily reachable information about level curves in our data, exhibit 7 does a quite reasonable job of portraying the dependence of the value of double flog upon the two cuts -- one horizontal and one vertical -- that combine to determine it.

* going on *

Exhibit 8 collects some problems, projects, and major projects that could be undertaken as part of going on with this data.

29yC. <u>"Smooth curving" a fit</u>

We now have a reasonable picture of the dependence of our double flog on its two constituent cuts. Even though we believe that it is telling us about the errors of census enumeration rather than about the population of England and Wales, we may want to try to fit it as well as we can with some moderately simple expression, perhaps

exhibit 8 of chapter 29y -- married couples

Problems and projects open

A) The OTHER set of CROSS-SECTIONS -- a project

Problem 1. Construct the analog of exhibit 5 for the cross-sections for each wife's age cut. (We naturally preserve the horizontal scale of exhibit 2 and plot "height" vertically.)

Problem 2. Construct from this the analog of exhibit 6, and, using this analog only, construct the analog of exhibit 7.

Problem 3. Combine exhibit 6 and the analog just mentioned on the same sheet of tracing paper. Draw an analog of exhibit 7 using both sets of intersections.

Problem 4. Trace back -- to exhibit 5 or its analog -- the changes needed to bring intersections -- in exhibit 6 or its analog -- onto the smooth curves that ultimately arise from problem 3.

B) A LITTLE MORE SUPPORT -- a project

Problem 5. Using the following data -- of the form (count for wives 20-24, count for wives 25-29) -- (1342, 7), (219078, 45284), 318865, 519689), (84583, 435381), (17064, 123387), (4231, 30130), (1261, 8269), (441, 2601), (160, 923), (95, 442), (32, 215), (21, 61), (8, 23), (5, 23), (4, 8), (2, 2), (5, 6) for 5-year intervals of husband's age supply the double flogs for wife's age cut b.

Problem 6. Given this additional row in the basic table of double flogs, redo exhibits 5, 6 and 7 as necessary.

Problem 7. Given this additional row in the basic table, redo as much of problems 1, 2, 3 and 4 as necessary.

C) MUCH MORE DETAIL -- a major project

Problem 8. Look up the source of exhibit 1 and obtain the counts for 5-year intervals in both wife's and husband's ages.

Problems 9-13. Find the double flogs for all possible cuts of the data of problem 8. Set aside rows and columns with flogs of size greater than 3 (beyond which there are less than 1 couple in 1000).

exhibit 8 (cont'd.)

Problems 14-16. Construct the analogs of exhibit 5 using all the double flogs not set aside. (With the increased number of cross-sections it may pay to divide these cross-sections -- with suitable overlaps -- into 2 or 3 plots.

Problem 17. Construct the analog of exhibit 6.

Problem 18. Construct the analog of exhibit 7.

Problems 19-23. Do the same, with wife's age interchanged with husband's age (as in problems 1 to 4).

D) PARALLEL DATA -- a project

Problem 24. Look up a smaller set of data for some other data in Britain or for some other country. Collect the counts corresponding to panel A of exhibit 1.

Problem 25. Calculate the corresponding double flogs.

Problem 26. Construct the analog of exhibit 5.

Problem 27. Construct the analog of exhibits 6 and 7.

Problems 28-31. Carry out the analogs of problems 1 to 4.

f(sum) plus g(diff)

where f() and g() are reasonably smooth functions and, as in section 29xF,

sum = (column flog) plus (row flog),

diff = (column flog) minus (row flog).

A look at exhibit 7 is likely to convince us that we are not likely to do very well if we restrict these functions to be easily manageable polynomials. We need to have a way to handle -- at least roughly -- more general smooth functions. Exhibit 9 shows the details of one way of starting to do this.

We begin by arranging the data pairs according to the value of the carrier, smooth functions of which we are to contemplate as possible fits. Panel A shows the values of carrier and response -- the latter actually a residual in this example -- and the finding of a median for each batch of 5 consecutive responses. Panel B compresses this 1st step and details the 2nd, in which medians are found for each batch of three consecutive results of the 1st step -- for each batch of three consecutive running medians of batches of five consecutive responses. Panel C compresses the whole calculation into the form we shall usually report.

exhibit 9 of chapter 29y -- running medians

A detailed example of the smoothing-by-running median calculation for a short sequence of residuals

A) DETAILS

ordered carrier	residual	blocks of five residuals and their median												
-316	5	15												
-256	21	21		21										
-216	-33	-33	21	-33		-33								
-211	21	21		21	21	21		21						
-168	25	25		25		25	-16	25		25				
-164	-16			-16		-16		-16	-1	-16		26		
-150	-26	-26				-26		-26		-26	-7	26		-26
-121	-1	-1						-1		-1		-1	-7	-1
-114	-7	-7	-7							-7		-7		-7
-98	-16	-16										-16		
-90	-7	-7												

B) STEP 1 SUMMARIZED, STEP 2 DETAILED

ordered carrier	residual	running median	blocks of three running medians and their medians								
-316	15										
-256	21										
-216	-33	21		21							
-211	21	21		21	21	21					
-168	25	-16		-16		-16	-1	-16			
-164	-16	-1				-1		-1	-7	-1	
-150	-26	-7		-7				-7		-7	-7
-124	-1	-7		-7	-7					-7	
-114	-7	-7		7							
-98	-16										
-90	-7										

C) STANDARD SUMMARIZED FORM

ordered carrier	residual	rM5M3
-316	15	
-256	21	
-216	-33	21
-211	21	21
-168	25	-16,-1
-164	-16	-1,-7
-150	-26	-7,
-121	-1	-7,
-114	-7	-7,
-98	-16	
-90	-7	

Note: "rM5M3" stands for running medians of 5 followed, after a comma, by running medians of 3, where this changes anything.

D) DATA SOURCE

Part of one of the later exhibits in this section.

In practice, it is often worthwhile to write the carrier and response down neatly (with the lines equally spaced), and then to tear a square notch in a piece of paper of the right size to allow us to use it to look at only 5 adjacent entries in a column. We can then rapidly slide this notch along, finding one median of five mentally in each position.

Once we have found the values of "rM5M3", the next step is to plot the results and look hard. There are now two possibilities:

- we can see that some simple dependence -- usually linear or quadratic -- will fit the plot fairly well. When this is so we go ahead and make the fit in some more or less reasonable way. Or

- it seems that any fairly smooth and fairly closely fitting curve will not be so simple. When this happens we make an eye fit by drawing in a suitable curve, balancing quality of fit against simplicity of shape in some intuitive way.

Often we find it desirable to go through the procedure twice both in order to be sure of the result and in order to be able to be clear about the first fit. The whole sequence of steps would then run as follows:

- order by carrier value; find "rM5M3" for response.

- plot "rM5M3" against carrier; fit a curve.

- take fit values off the curve at carrier values and calculate residuals.

- write down residuals against ordered carrier values and find new "rM5M3" values based on residuals.

- plot new "rM5M3" values, and make new fit -- when and to what extent called for. (The new fit may well be zero except for one or more intervals where the first fit needs adjustment.)

- take new fit values from new curve and calculate new residuals.

- take initial fit PLUS new fit as the fit found and the new residuals as the corresponding residuals.

Exhibits 10 and 11 show the twice-through process applied to the number of Presidential electors, 1828-1928, who voted for the Democratic candidate. No algebraically simple function can be expected to deal effectively with the Civil War period; we have to use eye fits. ("Before 1852" and "after 1880", as problem 1 shows, fit together surprisingly well.)

* end values *

In dealing with the extreme ends of the sequences, we have adopted the following rules of thumb:

- the 3rd from either end is smoothed normally by a running median of 5.

exhibit 10 of chapter 29y -- Democratic electors

Smooth curving the number of Presidential electors voting Democratic, 1828-1928, by running medians
(see exhibit 11 for plots and eye fits)

A) The COMPUTATION

Year	Votes	rM5M3	loose eye fit	1 res	rM5M3	close eye fit	2 res	total fit
1828	178	(178)	175	3	(3)	3	00	178
1832	219	178	169	50	4	4	45	173
1836	170	170	166	4	4	4	0	170
1840	60	170	162	-102	4	4	-106	166
1844	170	170	159	11	4	7	4	166
1848	127	170	156	-29	11	10	-39	166
1852	254	170	150	104	11	10	94	160
1856	174	127	127	47	9	10	37	137
1860	12	21	21	-9	9	8	-17	29
1864	21	12	12	9	-9	-3	12	9
1868	0	12	12	-12	-9	-11	-1	1
1872	0	21	21	-21	-12	-12	-9	9
1876	184	155	156	28	-12	-11	39	145
1880	155	168	167	-12	-10	-8	-4	159
1884	219	184,176	177	42	28,1	-2	44	175
1888	168	176	178	-10	1	1	-11	179
1892	277	176	176	101	1	1	101	177
1896	176	168	175	1	-10,-1	0	1	175
1900	155	162	171	-16	-1	0	-16	171
1904	140	162	168	-28	-1	0	-28	168
1908	162	162	163	-1	-1	-1	0	162
1912	435	162	157	278	-1	-1	279	156
1916	277	162	150	127	-1	-2	129	148
1920	127	136	145	-28	-2	-2	-26	143
1924	136	136	138	-2	-2	-2	0	136
1928	87	(87)	132	-45	(-45)	-2	-43	130

(cont'd.)

exhibit 10 (cont'd.)

Examples:

170 = median of 178,219,170,60,170's ...; 136 = median of 435,272,127,136,87.

",178" = median of 178,219,170 (as just found); ",136" = median of 87,136,136 (as just found).

",(178)" = copy of 178; ",(87)" = copy of 87.

6 = 178 - 172; 49 = 219 - 170; ... -2 = 136 - 138; -45 = 87 - 132.

6 = median of 6,49,6,-102,11; etc.

0 = 6 - 6; 43 = 49 - 6; ... 0 = (-2) - (-2); -43 = (-45) - (-2).

178 = 172 + 6; 176 = 170 + 6; ... 136 = 138 + -2; 130 = 132 + -2.

B) PROBLEMS

1. Plot the second residuals "2 res" against date; repeat at an expanded scale fitting 1860 to 1872.

2. Plot the final fit against date.

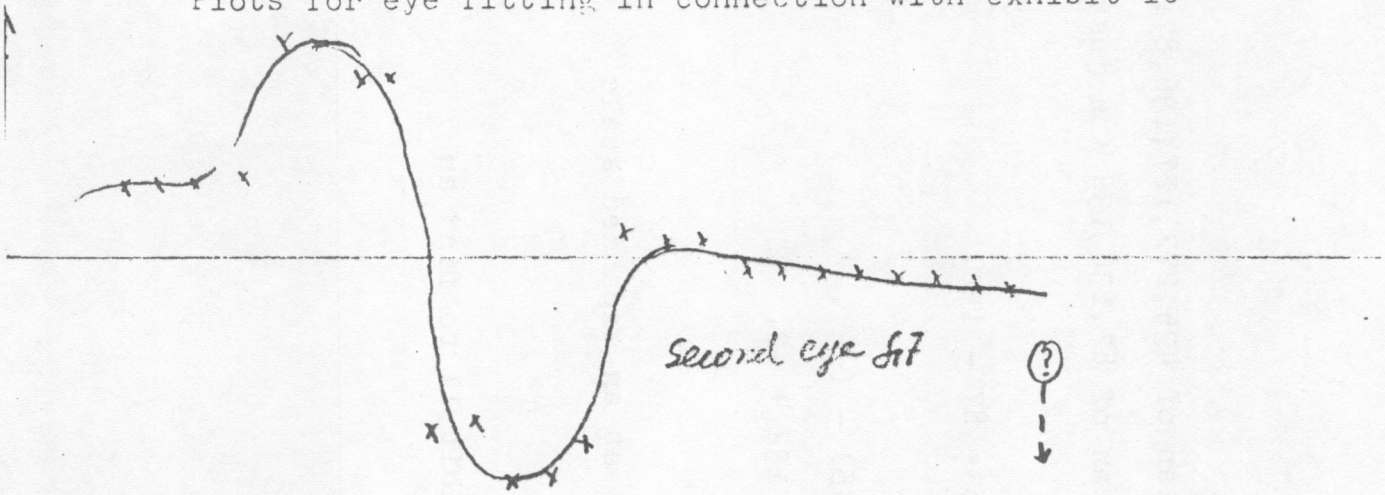
3. Use the dashed lone eye fit in the lower panel of exhibit 11 to find an alternate analysis.

4. Do problems 1 and 2 for this alternate analysis.

29y-22

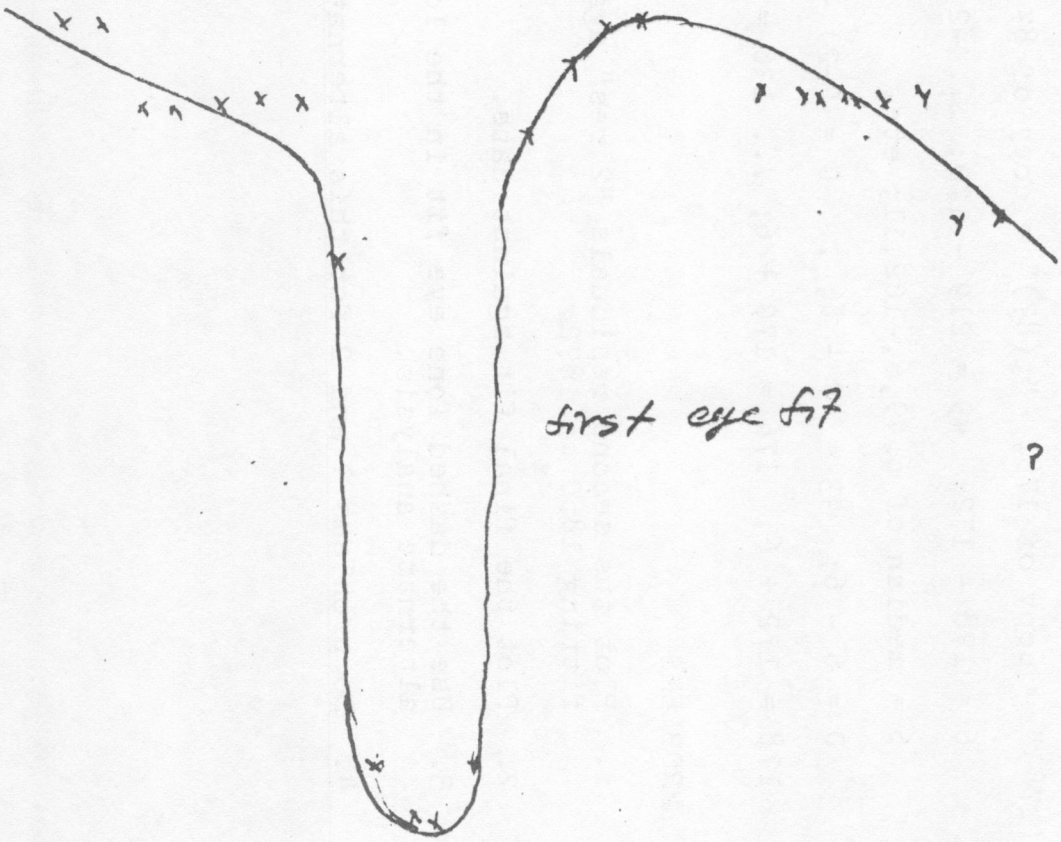
exhibit 11 of chapter 29y -- Democratic electors

Plots for eye fitting in connection with exhibit 10



1850
|
1850

1900
|
1900



- the next to either end is the median of the end three values.

- the end entry is repeated, in parentheses, in the column for medians of 5. (It is plotted with a "?" NOT and "×".)

At the upper end of the first smoothing of exhibit 10, for instance, we have, successively

178	(178)
219	178
170	170'
50	
170	
...	

where (178) repeats 178, 178 is the median of 170, 219 and 178 and 170 is the median of 170, 60, 170, 219, 178. (The reader is advised to check through the details at the foot of the same column.

29yD. Application to married couples

Exhibit 12 begins the application of this technique to the "married pairs" double flogs. Panel A sets out the values of

diff = column flog MINUS row flog

and for repeated use

sum = column flog PLUS row flog

Panel B of exhibit 12 and the two panels of exhibit 13 carry out one round of (twice-through) smooth curving. It is clear from the upper plot of exhibit 13 that we need eye fits rather than simple polynomials. After this round the 2nd residuals have midspread .27h and 8thspread .52. The last column gives the values of the next carrier. They are useful in writing down the beginning of the next round.

The next round smooth curves the 2nd residuals against "sum". Exhibit 14 and 15 have the details. As usual we fit loosely the first time through, but more closely the second time. The last column again gives the values of the next carrier, now "diff". The 4th residuals now have midspread .15h and 8thspread .34h, only about two-thirds their previous sizes.

It would be much too naive to stop here. Each fit of a smooth function of one carrier alters the problem of fitting the other -- almost surely toward greater opportunity. We ought to expect to go round at least once more with each carrier being rather more careful.

Exhibits 16 and 17 go once through again against "diff", starting now with 4th residuals. In the upper plot of exhibit 17 we do the loose fit by cutting across loops A

exhibit 12 of chapter 29y -- married couples

1st round of smooth curving -- relation to "diff"

A) The values of DIFF and SUM -- unit 0.01

	diff							sum						
	b	c	d	e	f	g	h	b	c	d	e	f	g	h
a	80	146	206	251	294	348	440	-382	-316	-256	-211	-168	-114	-22
c	-88	-20	40	85	128	182	274	-216	-150	-90	-45	-2	52	144
d	-138	-72c	-12d	33e	76	130	222	-164	-98	-38	7	508	104	196
e	-181	-115c	-55d	-10	338	87	189	-121	-55	5	50e	93	147	239
f	-226	-160c	-100	-55e	-12f	42	134	-76	-10	50d	95	138	192	284
g	-286	-220	-160d	-115e	-72f	-18	74	-16	50c	110	155	198	252	344
h	-387	-321	-261	-216	-173	-119	-27	85	151	211	256	299	353	445

(cont'd.)

29y-26

exhibit 12 (cont'd.)

B) SMOOTH CURVING against DIFF

diff	fflog	rM5M3	eye fit	1 res	rM5M3	fit	2 res	fit so far	(sum)
-387	209	(209)	205	4	(4)	6	-2	211	(85)
-321	246	246	240	6	6	5	1	245	(151)
-286	284	248	260	24	6	4	20	264	(-16)
-261	248	284	276	-28	20	4	-32	280	(211)
-226	314	284	294	20	20	3	17	297	(-76)
-220	319	314	296	23	5	3	20	299	(506)
-216	272	314	300	-28	5	3	-31	303	(256)
-181	319	319,317	314	5	5	9	2	317	(-121)
-173	267	317	315	-48	1	2	-50	317	(299)
-160c	330	317	316	14	1	2	12	318	(-70)
-160d	317	299,309	316	1	-15	2	-1	318	(110)
-138	299	316,309	314	-15	1	1	-16	315	(-164)
-119	252	309	311	-50	0		-59	311	(353)
-115c	346	309	309	9	0		7	309	(-55)
-115e	309	309	309	0	0		0	309	(155)
-100	318	309	361	17	-1		17	301	(50d)
-86	248	280	289	-61	-1		-41	289	(-216)
-72c	280	280	281	-1	-1		-1	281	(-98)
-72f	271	278	281	-10	-1		-10	281	(198)
-55d	288	278	266	22	12		22	266	(5)
-55e	278	271	266	12	12		12	268	(95)
-27	247	247	234	13	0		13	234	(445)
-20	203	223	226	-23	0		-23	226	(-150)
-18	223	222,223	223	0	0		0	223	(252)
-12d	214	222,222	221	-7	0		-7	221	(-38)
-12f	222	222	221	1	1		1	221	(138)
-10	223	214	220	3	1		3	226	(50c)
33e	210	210	208	2	2		2	208	(7)
33f	205	210	208	-3	2		-3	208	(93)
40	210	205	208	2	-3		2	208	(-90)
42	196	205	208	-12	-3		-12	208	(192)
74	201	201	212	-11	-11		-11	212	(344)
76	235	201	213	22	-11		22	213	(50f)
80	194	214	215	-21	-2		-21	215	(-382)
85	250	235	215	35	22		35	215	(-45)
87	214	250	216	-2	26		-2	216	(147)
128	279	250	225	54	26		54	225	(-2)
130	252	214,250	226	26	-2,226		26	226	(104)
134	197	252,219	226	-29	26,-2		-29	226	(284)
146	213	219	230	-17	-17		-17	230	(-316)
182	295	219	280	55	-17		55	240	(52)
189	219	243	242	-23	-6		-23	242	(239)
206	250	250	245	5	5,-6		5	245	(-256)
222	243	250	249	-6	-6,5		-6	249	(196)
257	270	251	256	14	5,0		14	256	(-211)
274	251	270	261	-10	0		-10	261	(144)
294	291	270	266	25	0		25	266	(-168)
348	276	276	271	5	5		5	272	(-114)
440	231	(231)	276	-45	(-45)		-45	276	(-22)

(median) (1)

(Hinges) (-11h and 13h) (-11 and 13h)

(8ths) (-28 and 23) (-30 and 22)

clarity (diff only) 4.0 = $\frac{318-211}{27h}$

C) PROBLEMS

1. Plot the final fit against "diff".

2. Try using the dashed eye fit in the lower panel of exhibit 13.

3. Do problem 1 completing the alternative version of panel B for the alternative fit.

exhibit 13 of chapter y -- marrie ouples

Plot of eye fits connected with exhibit 12

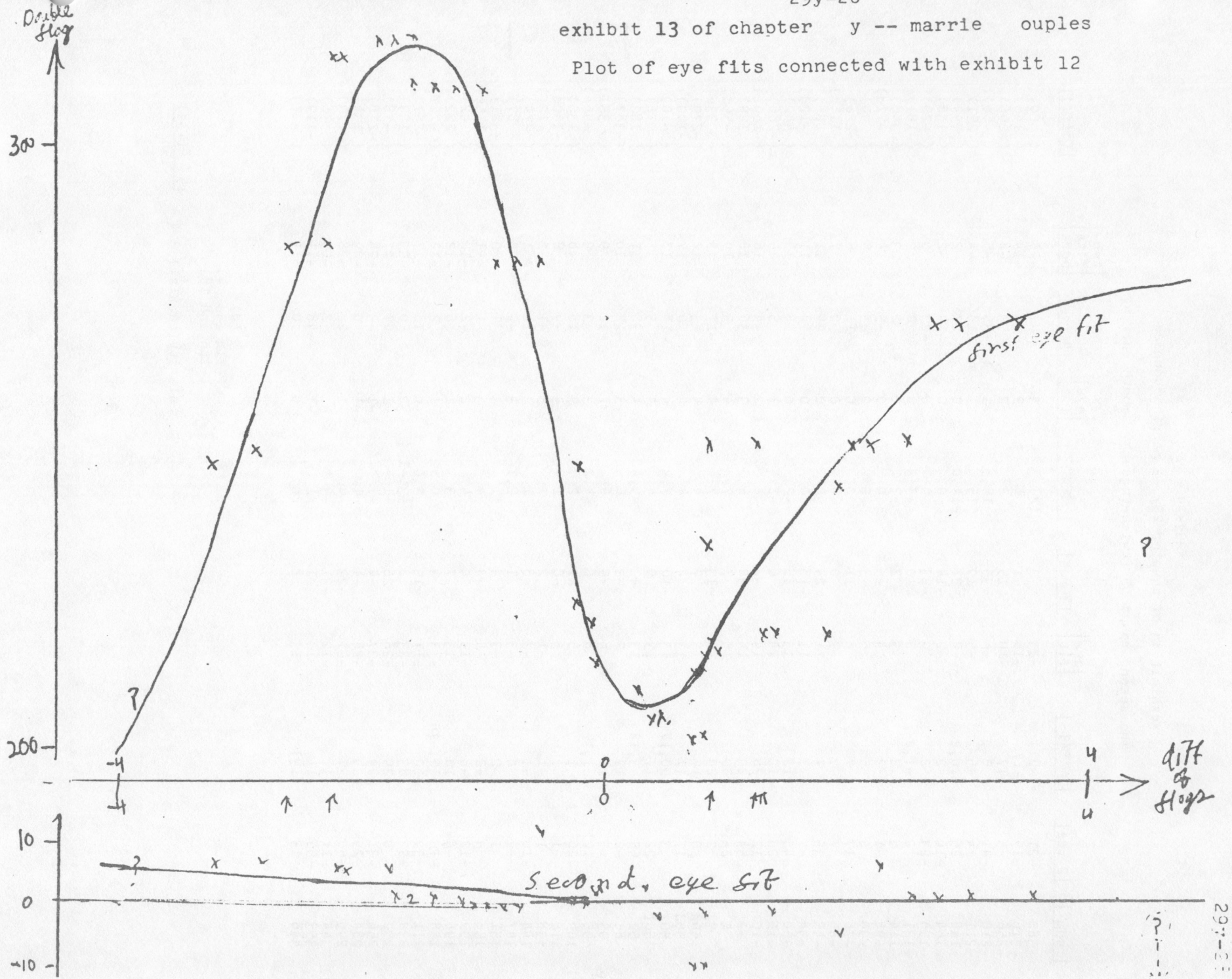


exhibit 14 of chapter 29y -- married couples

The second round of (twice-through) smooth curving

sum	2 res	rM5M3	eye fit	3 res	rM5M3	eye fit	4 res	two fits here	(diff)
-382	-21	(-21)	-23	2	(2)	3	-1	-20	(80)
-316	-17	-17	-16	-1	2	2	-3	-14	(146)
-256	5	-17	-10	15	2	2	13	-8	(206)
-216	-41	5	-6	-35	15	1	-36	-5	(-86).
-211	14	5	-5	19	15	1	18	-4	(251)
-168	25	-16,2	-2	27	-14,0	0	27	-2	(294)
-164	-16	2	-2	-14	0	0	-14	-2	(-138)
-150	-23	2	0	-23	0	0	-23	0	(-20)
-121	2	-1,2	2	0	-5,3	0	0	2	(-181).
-114	5	2	3	2	-3	0	2	3	(348)
-98	-1	2	4	-5	0	0	-5	4	(-72c)
-90	2	5	5	-3	0	0	-3	5	(40)
-76	17	7	6	11	0	0	11	6	(-226)
-55	7	7	7	0	0	0	0	7	(-115c)
-45	35	7	11	24	0	0	24	11	(85)
-38	-7	7	11	-18	0	0	-18	11	(-12d)
-22	-45	12	13	-58	-2	0	-58	13	(440)
-16	20	12	14	6	-2	1	5	15	(-286)
-10	12	20	14	-2	6	2	-4	16	(-160c)
-2	54	20	15	39	6	3	36	18	(128)
5	22	20	15	7	7	4	3	19	(-55d)
7	2	20	15	-13	7	4	-17	19	(33e)
50c	20	17	13	7	4	4	3	17	(-220)
50d	17	17	13	4	4	4	0	17	(-100)
50e	3	20,17	13	-10	7,4	4	-14	17	(-10)
50f	22	17	13	9	4	4	5	17	(76)
52	55	3,12	13	42	-9,4	4	38	17	(182)
85	-2	12	7	-9	6	3	-12	10	(-387)
93	-3	12	6	-9	6	7	-12	9	(33f)
95	12	-1,1	6	6	-4,2	3	3	9	(-55e)
104	26	1	5	21	2	3	18	8	(130)
110	-1	1	5	-6	2	3	-9	8	(-160d)
138	1	-1	-1	2	1	3	-1	2	(-12f)
144	-10	-1	-2	-8	1	3	-11	1	(274)
147	-2	0	-3	1	2,1	3	-2	0	(87)
151	1	-2	-3	4	1	3	1	0	(-321)
155	0	-2	-4	4	4	3	1	-1	(-115e)
192	-12	-6	-12	0	4	2	-2	-10	(42)
196	-6	-10	-13	7	3	1	6	-12	(222)
198	-10	-12,-10	-13	3	0.	1	2	-12	(-72f)
211	-32	-10	-17	-15	3,0	0	-16	-17	(-261)
239	-23	-23	-23	0	-3,-1	0	0	-23	(189)
252	0	-29	-26	26	-1	-1	27	-27	(-18)
256	-31	-29	-27	-4	-1	-1	-3	-28	(-218)
284	-29	-29	-28	-1	-1	-1	0	-29	(134)
299	-50	-27	-27	-23	-4	-1	-22	-28	(-173)
344	-11	-11	-12	-31	-3	-1	2	-13	(76)
353	-59	-11	-10	-1	1	-1	-48	-11	(-119)
445	13	(13)	13	0	(0)	0	0	13	(-27)

(0)
(-11h and 4)
(-16h and 18)

clarity(partial) 3.4 = $\frac{19 - (-34)}{15h}$

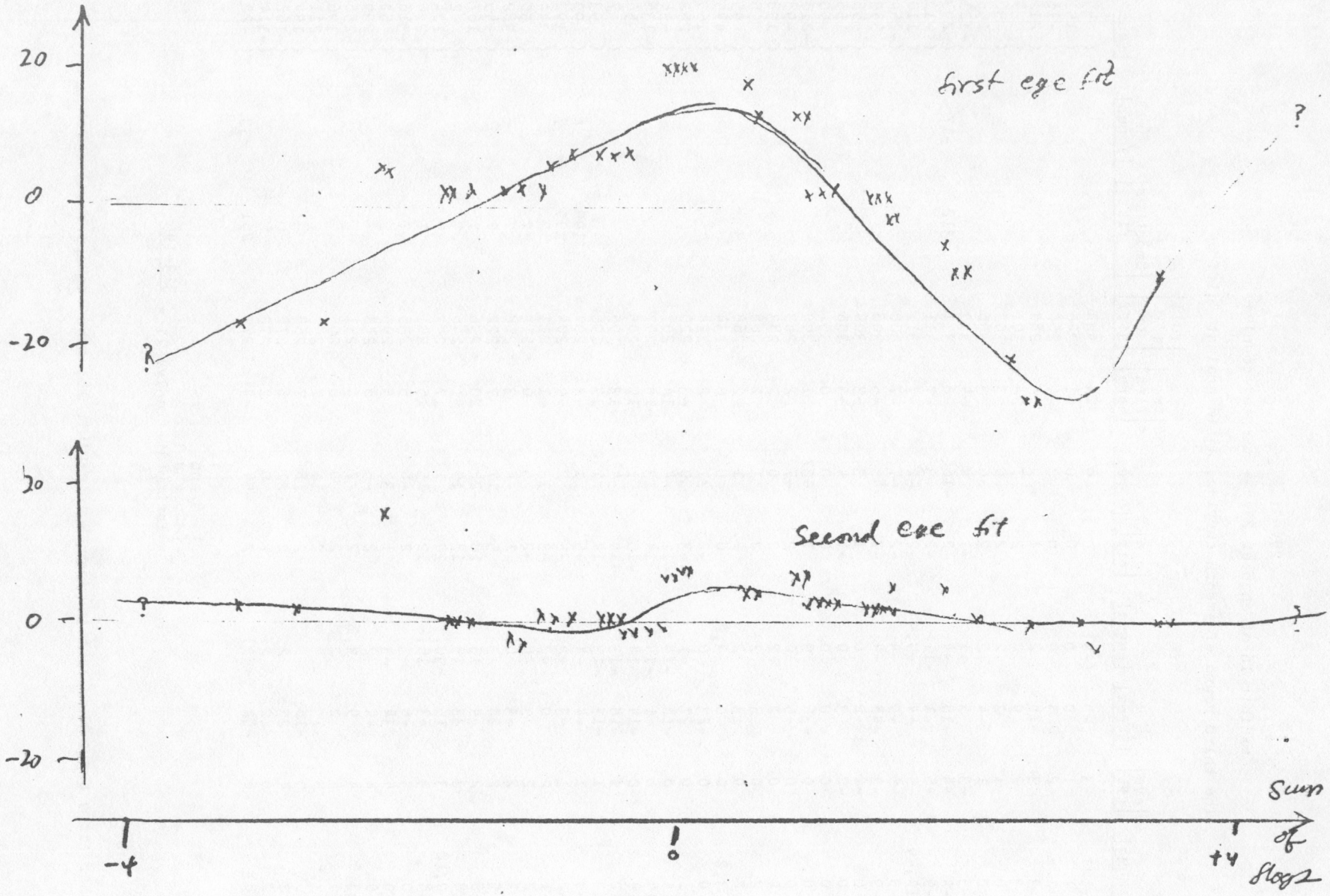


exhibit 15 of chapter 29y -- married couples

Plots for eye fits connected with exhibit 14

exhibit 16 of chapter 29y -- married couples

The third round of smooth curving, again against "diff"

'f	4 res	rM5M3	eye fit	5 res	rM5M3	eye fit	6 res	two fits here	full fit on diff	Smoother fit	7 res	(sum)
-387	-12	(-12)	-3	-9	(-9)	0	-9	-3	208			(85)
-321	1	1	1	0	0	0	0	1	246			(151)
-286	5	1	2	3	0	0	3	2	266			(-16)
-261	-15	3	3	-18	2	1	-19	4	284			(211)
-226	11	3	2	9	2	2	7	4	301			(-7c)
-220	3	0	1	2	2	2	0	3	302			(50e)
-216	-3	0	1	-4	2	2	-6	3	306			(256)
-181	0	-3	-3	3	1	-1	4	-4	313			(-121)
-173	-22	-4	-4	-18	-4	-2	-16	-6	311	312	-17	(299)
50c	-4	-9	-5	1	-11	-2	3	-7	311			(-10)
50d	-9	-14,-9	-5	-4	-9,-4	-2	-2	-7	311			(110)
-138	-14	-9	-5	-9	-4	-2	-7	-7	308			(-164)
-119	-48	-9	-3	-45	-4	-2	-43	-5	306			(353)
-115c	0	0	-2	2	0	-1	3	-3	306			(-55)
-115e	1	0	-2	3	0	-1	4	-3	306			(155)
-100	0	0	0	0	0	0	0	0	351			(50d)
-86	-36	0	0	-36	0	1	-37	1	290			(-216)
-72c	-5	0	0	-5	0	2	-7	2	283			(-98)
-72f	2	2	0	2	2	2	0	2	283			(198)
-55d	3	2	0	3	2	2	1	2	268			(5)
-55e	3	2	0	3	2	2	1	2	268			(95)
-27	0	3,2	0	0	3,2	0	0	0	234			(445)
-20	-23	0	0	-23	0	-2	-21	-2	224			(-150)
-18	27	-1	0	27	-1	-3	30	-3	220			(252)
-12d	-18	-14	0	-18	-14	-11	-4	-14	207			(-38)
-12f	-1	-14	0	-1	-14	-14	13	-14	207			(138)
-10	-14	-14	0	-14	-14	-15	1	-15	205			(50e)
33e	-17	-12	1	-18	-13	-14	-4	-13	195			(7)
33f	-12	-12	1	-13	-13	-14	1	-13	195			(93)
40	-3	-3	1	-4	-4	-5	1	-4	204	199	6	(-90)
42	-2	-2	1	-3	-3	-5	2	-4	204	199	7	(192)
74	2	-1	2	0	-3	-3	13	-1	211	?		(344)
76	5	2	2	3	0	-2	5	0	213	?		(50f)
80	-1	2	2	-3	0	1	-4	3	218	?		(-382)
85	24	5	2	22	3	5	17	7	222	?		(-45)
87	-2	18	2	-4	15	18	-22	20	236	?		(167)
128	36	18	3	33	15	18	15	21	246	?		(-2)
130	18	0,18	3	15	-30,15	14	1	17	243	?		(104)
134	0	18	3	-3	15,-3	12	-15	15	201	?		(236)
146	-3	0	3	-6	-3	2	-8	5	235	?		(-316)
182	38	0	4	34	-3	1	33	5	295			(52)
189	0	6	4	-4	2	2	-5	6	248			(239)
206	13	13,6	4	9	9,2	2	7	6	251			(-256)
22	6	6	4	2	2	2	0	6	255			(196)
51	18	13,6	5	13	9	2	11	7	263			(-211)
274	-11	0	5	-16	2	2	-18	7	268			(144)
294	27	2	4	23	2	2	21	6	272			(-168)
348	2	2	0	2	2	2	0	2	273			(-114)
440	-58	(-58)	-6	-52	(-52)	2	-54	-4	272	251	-33	(-22)

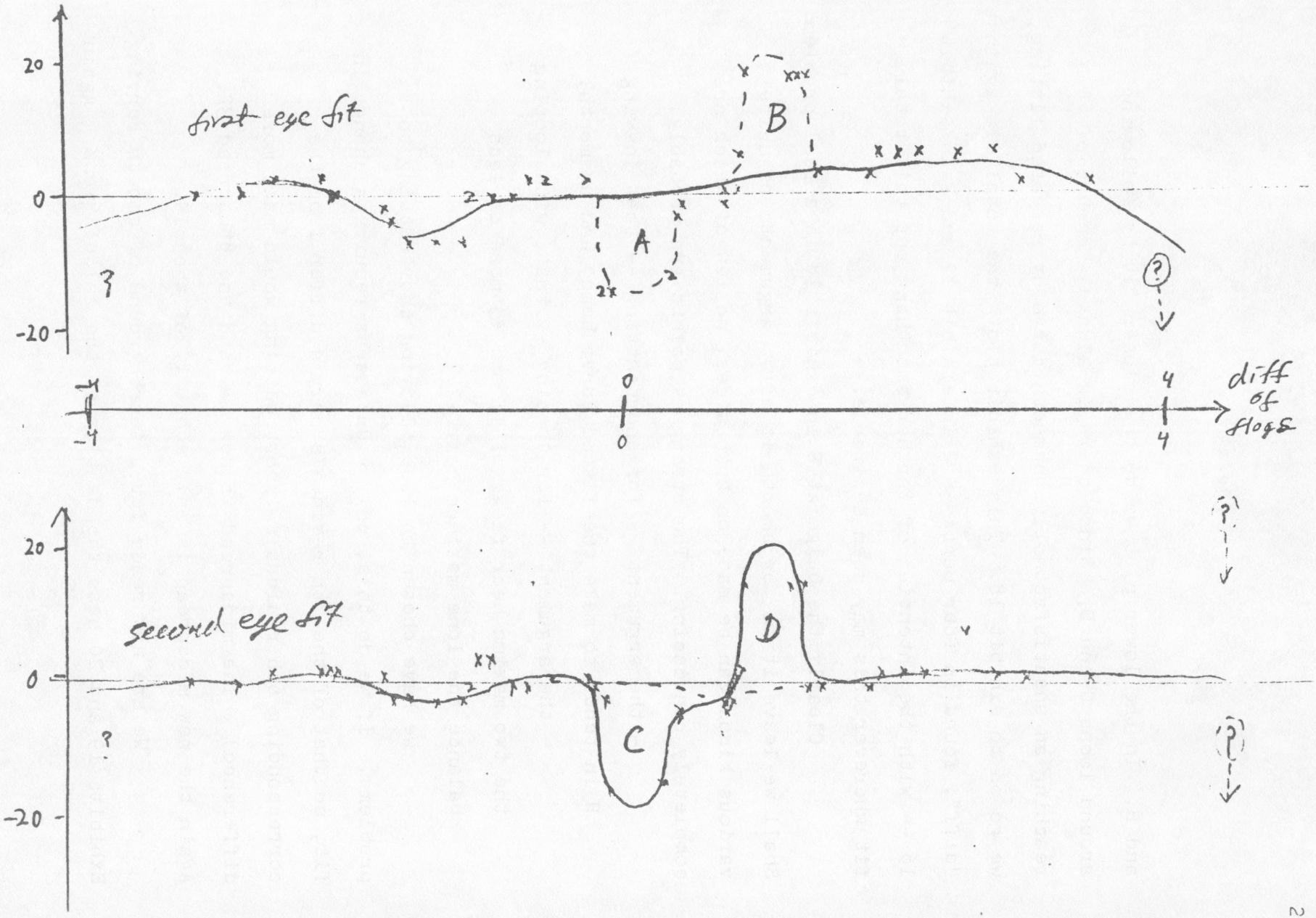
(0)
(-6 and 5)
(-20 and 12)
clarity(diff only) 11 = $\frac{313-195}{11}$

PROBLEM

Use as the smoother fit a curve short-cutting bump E in exhibit 18, find the appropriate 7th residuals.

Exhibit 17 of chapter 29y -- married couple

Plots for eye fits connected with exhibit 16



and B. In the lower plot we do the closer fit by following around loops C and D. Since there is always a chance of reaching an unsatisfactorily unsmooth curve by repeated fitting, we go on to exhibit 18. This exhibit plots the total fit against "diff", found in four parts -- 2 in exhibit 12 and 2 in exhibit 16 -- with the intention of further smoothing out this total fit wherever this may seem desirable.

Clearly the bulge at E in exhibit 18 is a real problem. Shall we leave it in, or smooth it off? Arguments of quite various kinds can be made on both sides, no few of which are completely convincing. The strongest points are probably:

- the argument -- for shortcutting -- that leaving H in tends to make the fitted curve bumpy and unsmooth.
- the argument -- for looping -- that, with looping the two maxima near flogs of +1 and +3 more or less balance the lone maximum near +2.

We have chosen to loop, leaving shortcutting to problems. If we do this, only 4 data sets require a change in fit, so that of the 7th residuals only 4 differ from the corresponding 6th residuals (short cutting would make more difference). The midspread is now .11 and the 8thspread 32. Again the new midspread is only about .7 of the old.

We are now ready for a fourth round of smooth curving. Exhibits 19 and 20 have the deatils of the calculations. Again

Exhibit 18 of chapter 29y -- married couple

Final smoothing of fit to diff
(reported in exhibit 16)

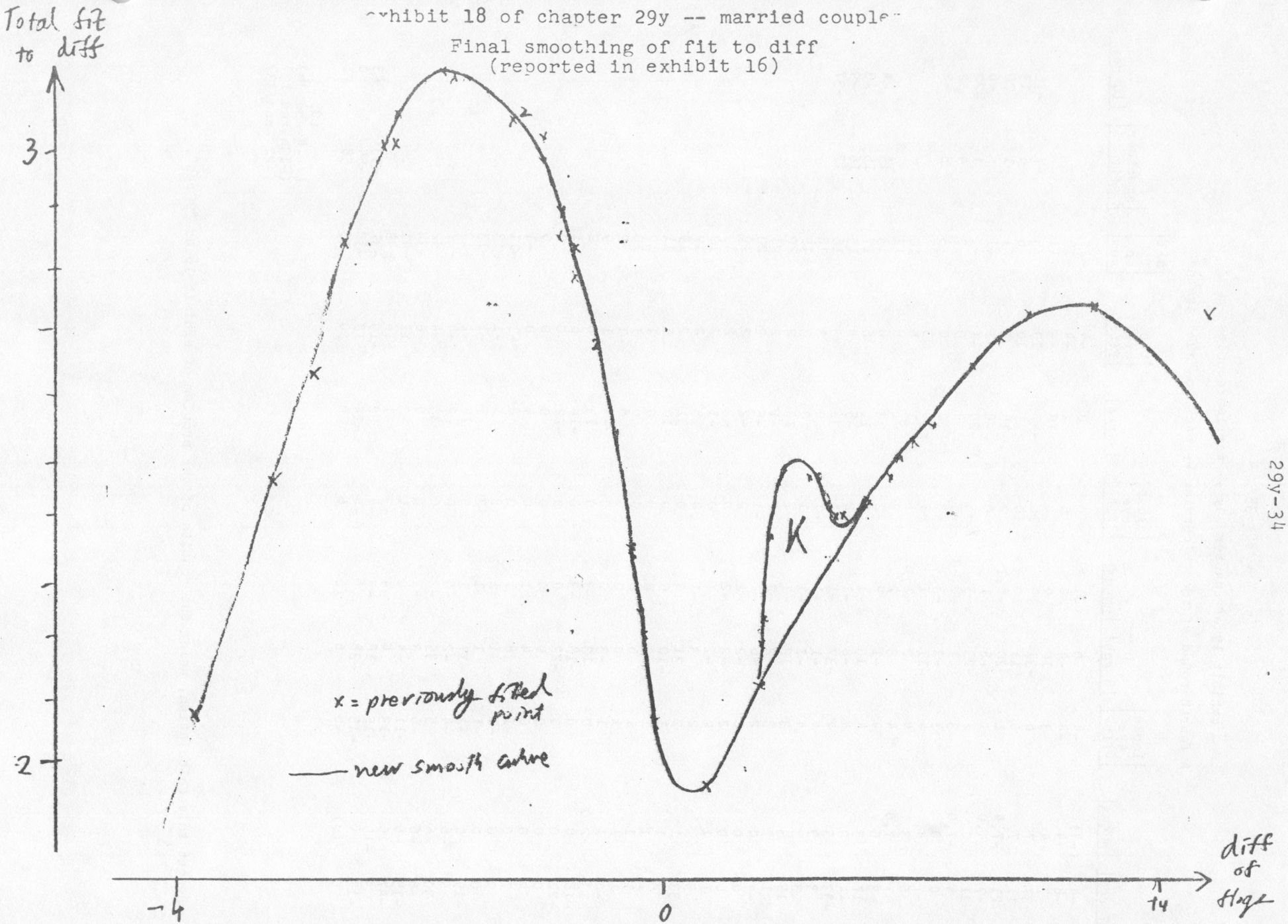


exhibit 19 of chapter 29y -- married couples

A fourth round of smooth curving -- back to "sum" again

sum	7 res	rM5M3	loose eye fit	8 res	rM5M3	close eye fit	9 res	two eye fits	all fit to sum	smoother fit	10 res
[illegible]	-4	(-4)	-4	0	(0)	0	0	-4	-24		
[illegible]	-8	-4	-4	-4	0	0	-4	-4	-18		
-256	7	-4	-3	10	0	0	10	-3	-11		
-216	-37	7	0	-37	10	10	-47	10	5	-7	-35
-211	11	7	0	11	10	10	1	10	6	-6	13
-168	21	-7,4	1	20	-8,1	2	18	3	1	-1	20
-164	-7	4,0	1	-8	1	0	-8	1	-1		-6
-150	-21	0	2	-23	-3	-3	-20	-1	-1	1	-22
-121	4	-7,0	3	1	-8,-3	-3	4	0	2	4	2
-114	0	0	3	-3	-3	-3	0	0	3	5	-2
-98	-7	4	4	-11	1	-1	-10	3	7		
[illegible]0	6	3	4	2	-1,1	1	1	5	10		
[illegible]6	7	7,6	4	3	2	3	1	6	12		
-55	3	6	4	-1	2	2	3	6	13	14	-4
-45	17	3	4	13	-1	2	11	6	17	15	13
-38	-4	3	4	-8	-1	0	-8	4	15	16	-9
-22	-33	3	4	-37	-1	-1	-36	3	16	17	-37
-16	3	3	4	-1	-1	-1	0	3	18		
-10	3	3	4	-1	-1	-1	0	3	19		
-2	15	3	3	12	-1	-1	13	2	20		
5	1	3	3	-2	-2	-1	-1	2	21		
7	-4	0	2	-6	-2	-1	-5	2	20		
50c	0	0	2	-2	-2	0	-2	2	19		
50d	0	0	2	-2	-2	0	-2	2	19		
50e	1	1	2	-1	-1	0	-1	2	19		
50f	5	1	2	3	-1	0	3	2	19		
52	33	1	2	31	0	0	31	2	19		
85	-9	1	1	-10	0	0	-10	1	11		
93	1	1	1	0	0	0	0	1	10		
95	1	1	1	0	0	0	0	1	10		
104	1	1	1	0	0	0	0	1	9		
110	-2	1	1	-3	0	0	-3	1	9		
138	13	-2,0	0	13	3	0	13	0	2		
144	-18	0	0	-18	-3	0	-18	6	1		
147	-22	0	0	-22	0	0	-22	0	0		
151	0	0	0	0	0	0	0	0	0		
155	4	0	0	4	1	0	4	0	-1		
192	7	0	-1	8	1	0	8	-1	-11		
196	0	0	-1	1	1	0	1	-1	-13		
198	0	0	-1	1	1	0	1	-1	-13		
211	-19	0	-2	-17	1	0	-17	-2	-19		
239	-6	-6	-4	-2	0	0	-2	-4	-27		
252	30	-6	-5	35	-2	0	35	-4	-32		
256	-6	-6	-6	0	-2	0	0	-6	-34		
284	-15	-6	-11	-4	0	-1	-3	-12	-41		
[illegible]99	-47	-15	-13	4	-4	-3	-1	-16	-44		
[illegible]44	3	-15	-15	18	-4	-4	22	-19	-32	-53	43
353	-13	0	-15	-28	6	-4	-24	-19	-30	-54	0
445	0	(0)	-6	6	(6)	6	0	0	13	-58	71

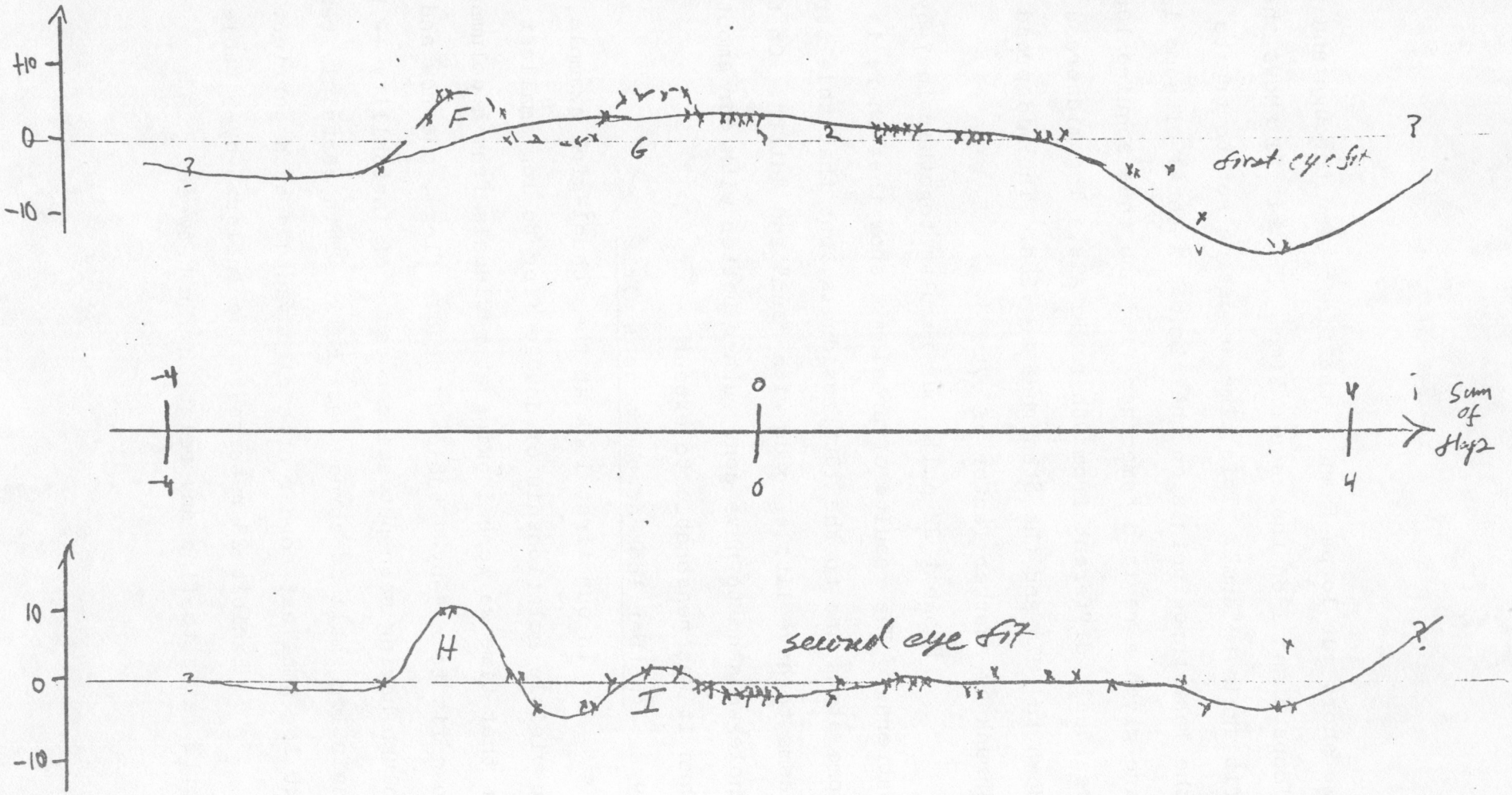
(0)
(-4 and 3h)
(-13h and 13)
(-22 and 33)

ROBLEM

Repeat this calculation, using the results of the problem of exhibit 16 as "7 res".

ibit 20 of chapter 29y -- married couple

Plots for eye fits in connection with exhibit 19



we short cut loops F and G the first time through and follow loops H and I on the second time. Exhibit 21 shows the total fit against "sum". This time we decide not to try to follow the last three points on the right. A total of some 13 points are given a revised "smoother fit" and thus acquire 10th residuals different from 9th residuals. The midspread is now down to .07h and the 8thspread to .26h. The midspread has shrunk by another factor of .7.

Exhibit 22 pulls our results together in 7-by-7 patterns. The results of problem 2 show that there is still some structure to the 10th residuals, but that this structure seems to have little to do with "sum" and "diff". On the whole, we seem to have done rather better with our smooth fit than it was reasonable to expect

29yE. Another look at the school girls

In our first look at the school-girl example, we saw no visible relationship of double flog to log smallest corner. At that time we were looking at residuals from a columns-plus-rows fit to a 4-by-4 table of double flogs, one row and one column having been put aside because of instability -- because their log smallest corner was small. This certainly tended to put log smallest corner at a disadvantage as a carrier.

Exhibit 23 returns to the fuller 5-by-5 table, but still fails to find any relationsihp between

nibit 21 of chapter 29y -- married coupl

Final smoothing of fit to sum
(reported in exhibit 19)

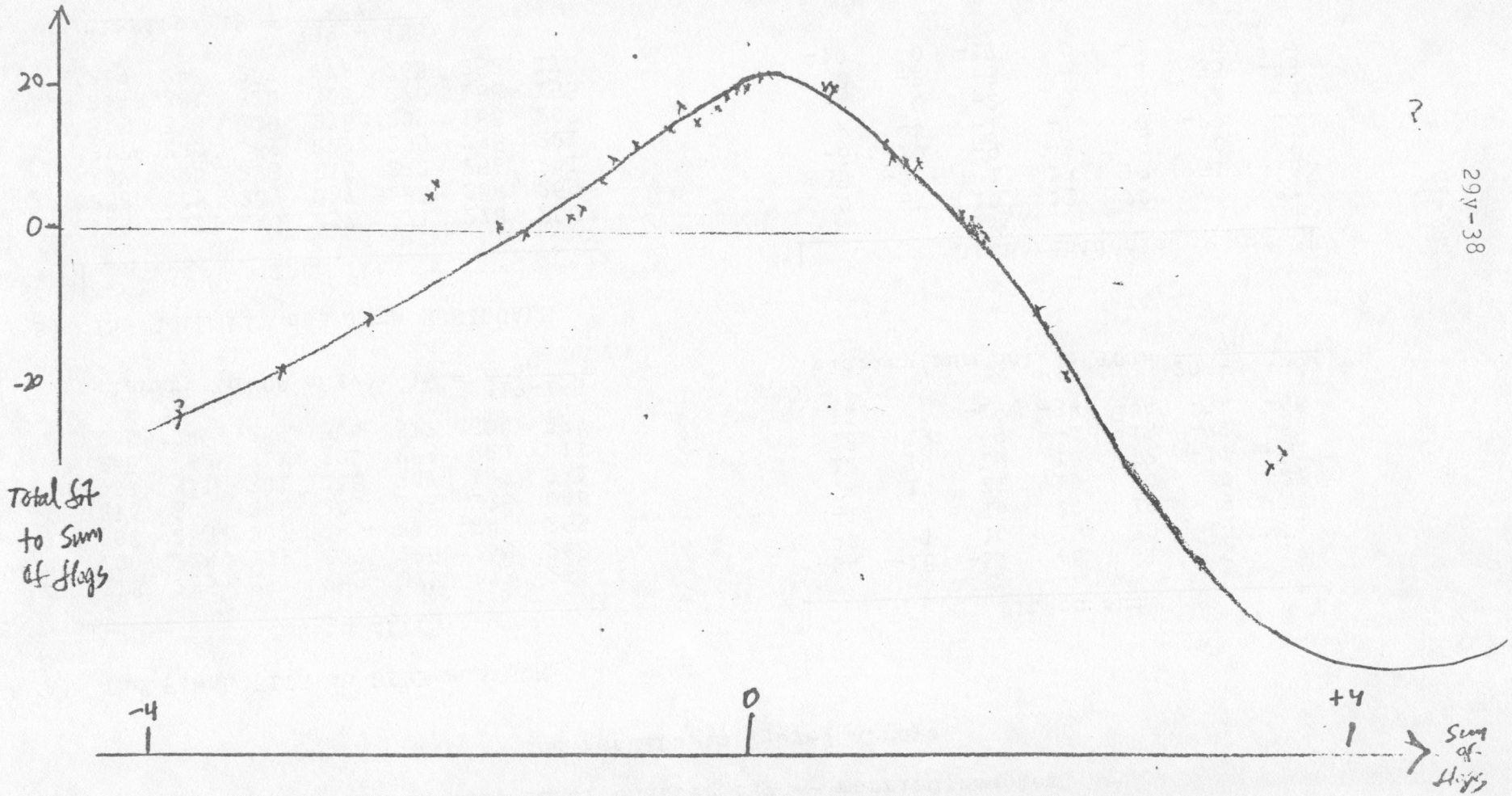


exhibit 22 of chapter 29y -- married couples

The results in 7-by-7 arrays

A) The FINAL FITS on DIFF and SUM

fit to diff

218	235	251	263	272	273	251
290	224	199	222	246	245	268
308	283	207	195	213	243	255
313	306	268	205	195	236	248
301	311	301	268	207	199	241
266	302	311	306	283	220	211
208	246	284	306	312	306	234

clarity (diff only): $16 = \frac{313-195}{7h}$

fit to sum

-24	-18	-11	-6	-1	5	17
-7	1	10	15	20	19	1
-1	7	16	20	19	9	-13
4	14	21	19	10	0	-27
12	19	19	10	2	-11	-41
18	19	9	-1	-13	-32	-53
11	0	-19	-34	-44	-54	-58

clarity(sum only): $10 = \frac{20 - (-58)}{7h}$

B) The 10th FIT and 10th RESIDUALS

10th fit

194	217	240	257	271	278	268
283	225	209	237	266	264	269
307	220	223	215	232	252	241
317	320	289	224	205	236	221
313	330	320	278	209	188	200
284	321	320	305	270	188	158
219	246	265	272	268	252	176

clarity: $19 = \frac{330 - 188}{7h}$

10th residuals

0	-4	10	13	20	-2	-37
-35	-22	1	13	13	31	-18
-8	-10	-9	-5	3	0	1
2	-4	-1	-1	0	-22	-2
1	0	-2	0	13	8	-3
0	-2	-3	4	1	35	-43
-10	0	-17	0	-1	0	7

Note: it is important to check that

fit PLUS residuals EQUALS data

(cont'd.)

exhibit 22 (cont'd.)

C) PROBLEMS

1. Add a 5th round of (twice-through) smooth curving. What happens to the midspread of residuals? To the 8thspread?

2. Convert the residuals to coded form and display 7-by-7.

D) PROJECT

Calculate values of fits to "diff" and "sum" at steps of .25 from -5 to 5. Use these values to calculate values of the smoothed fit on a 21-by-21 square from -2.50 to +2.50. Contour the result.

exhibit 23 of chapter 29y -- school girls

Another look at the school girl example at 5×5

A) The DOUBLE FLOGS and their COLUMN-PLUS-ROWS RESIDUALS

double flogs						residuals					eff
54	38	13	12	.		39	21	0	-11	-37	-34
41	43	33	28	.		0	0	-6	-21	0	-8
52	51	47	57	.		3	0	0	0	-6	0
34	52	53	73	.		-21	-5	0	10	33	6
46	58	64	66	.		-14	-4	6	-2	38	11
					eff	-2	0	-4	6	19	51

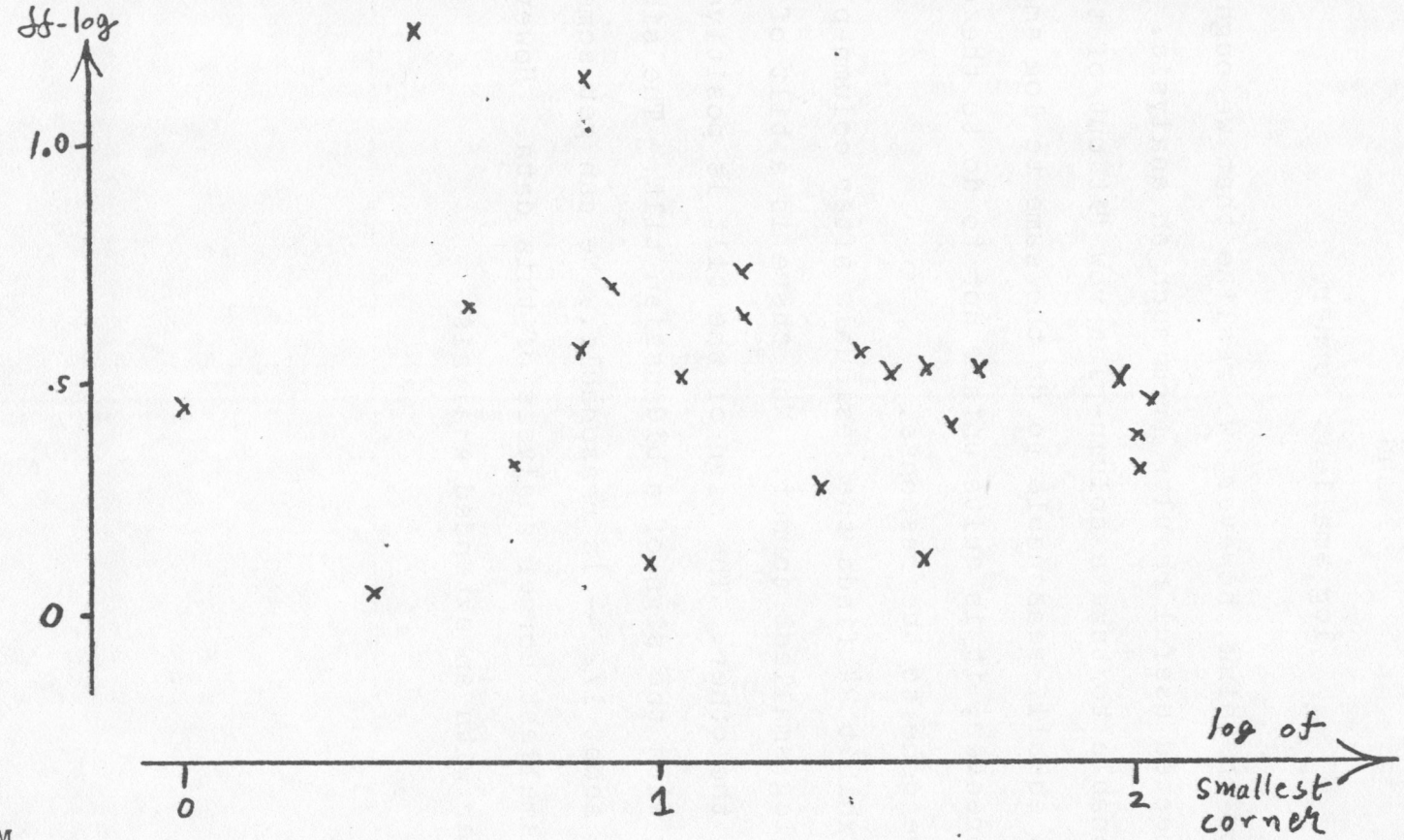
B) SMALLEST CORNERS, RAW and LOG -- logs in 0.01

smallest corners					log smallest corners				
38	102	35h	9h	2h	158	201	155	98	40
41	306h	101h	22	4	161	249	201	134	50
11	91h	106	26	8	104	196	203	142	90
5	30	47	15	7	70	148	167	118	85
1	7	15	4	3	0	85	118	60	48

(cont'd.)

exhibit 23 (cont'd.)

C) PLOT of ff-LOGS (from panel F of exhibit 4) against LOG SMALLEST COUNT



D) PROBLEM

Copy panel F of exhibit 4 and contour the pattern; do the same with log smallest count (panel A above); compare and describe the patterns.

29y-42

residuals from a column-plus-rows fit

and

log smallest corner

When we stop to think, however, we realize that we ought not to have expected useful results from such an analysis. If it is reasonable to take a column-plus-row fit out of the double flog, it is equally reasonable to do the same to log smallest corner. Moreover, it is quite unfair not to do to the carrier what we have done to the response.

Exhibit 24 finds the residuals after column-plus-row fitting to log smallest corner. Now there is a tilt of one residual on the other. The sign of the tilt is positive -- disagreeing with the sign of a biGaussian tilt. The size of the tilt -- about 1/2 -- is reasonable. We can get somewhere with a log smallest corner analysis of this data. However, we did better with an extended analysis.

exhibit 24 of chapter 29y -- school girls

Log smallest count after column-plus-row fit

A) TWO-WAY ANALYSIS of LOG SMALLEST CORNER -- three half-steps of median polish applied to panel B of exhibit 23, unit = 0.01

						eff
	90	41	0	-8	-14	-12
	47	5	0	-18	-40	34
	0	0	12	0	0	24
	-10	-24	0	0	19	0
	-22	-29	9	0	40	-58
eff	-14	78	49	0	-58	118

B) PLOT of RESIDUALS against RESIDUALS

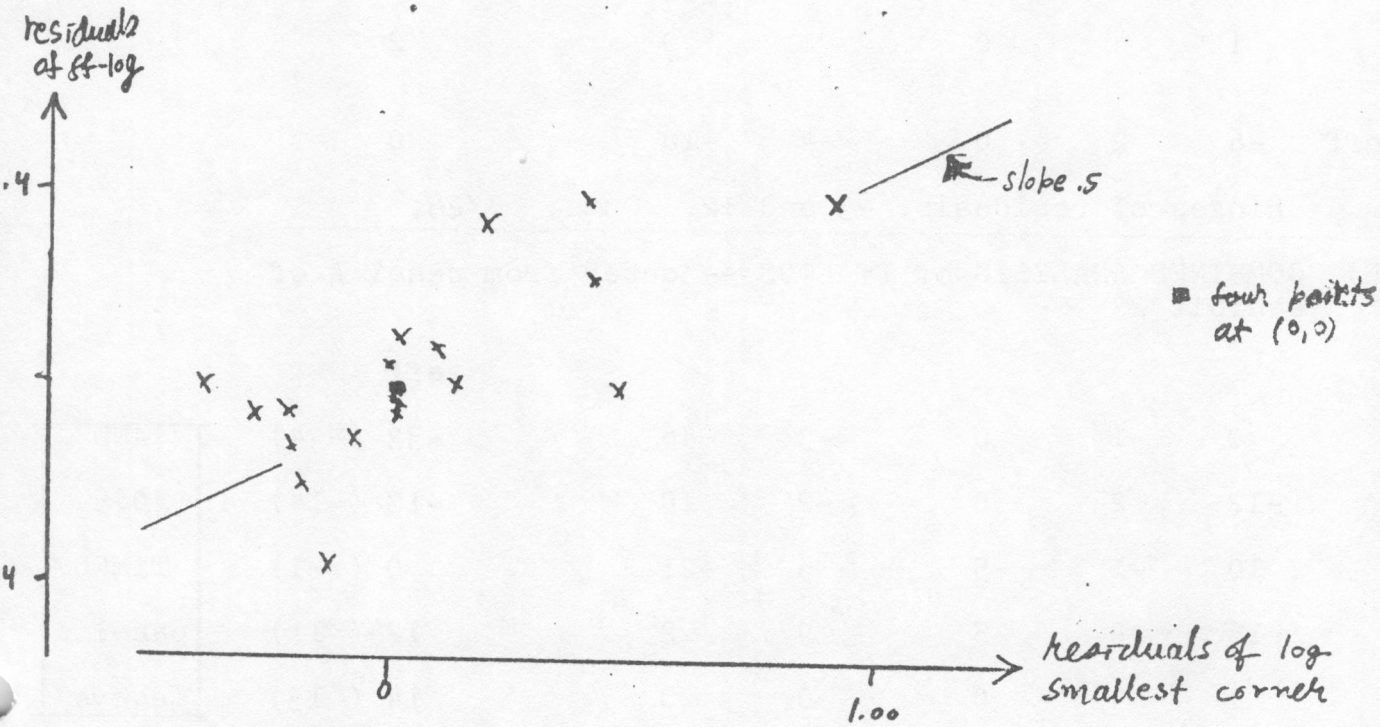


exhibit 24 (cont'd.)

C) NEW RESIDUALS -- panel A of exhibit 23 MINUS .5 times panel A

-6	1	0	-7	-30
-24	-2	-6	-12	20
3	-0	-6	0	-6
-16	7	0	10	23
-3	8	2	-2	18

D) INNER ANALYSIS of NEW RESIDUALS -- three half steps of median polish

						eff
	0	-1	0	-3	-46	0
	-12	2	0	-2	10	-6
	10	-1	-5	5	-21	-1
	-15	0	-5	9	2	5
	1	4	0	0	0	2
eff	-6	2	0	-4	-16	0

Hinges of residuals: -3 and +2. Outs: 7/25.

E) COMBINED ANALYSIS of ff-LOGS -- outer from panel A of exhibit 23

						eff	
	0	-1	0	-3	-46	-33 (-34)	AND
	-12	2	0	-2	10	-13 (-14)	0.5
	10	-1	-5	5	-21	0 (-1)	TIMES
	-15	0	-5	9	2	12 (11)	panel A
	1	4	0	0	0	14 (13)	above
eff	-10	2	-6	0	1	47	
	(-8)	(2)	(-4)	(2)	(3)	(51)	

Values in parentheses are sums of effects in the analyses of panel A of exhibit 23 and panel D above.

exhibit 24 (cont'd.)

F) PROBLEMS

1) Find the fit corresponding to panel E.

2) Subtract 0.5 times panel A from this fit, and check that the result is a possible column-plus-row fit.

3) Fit row effects and common only to the double flogs and the log smallest corners. Make the calculations parallel to panels B, C, D and E above from this starting point. Compare your results with those of panel E. Discuss the comparison.

****

29yH. <u>How far have we come?</u>

This chapter extends chapter 29x. In order to do this it introduces some new techniques, specifically:

- how to draw contours, starting with heights of the surface given at a regular grid of points.
- how to draw a smooth curve through (x,y) pairs when no simple function seems likely to work.

The process of contour drawing makes heavy use of the conveniences of tracing paper, starting with a family of cross-sections of the surface. These cross-sections are drawn through the heights at the grid points in each of several parallel sections, with considerable attention to the relationship of each cross-section to its neighbor. The intersections of these cross-sections with each of a family of level planes are marked, and then transferred to appropriate locations on a new sheet of tracing paper, where they serve to guide the drawing of the corresponding level curves or contours.

The process of fitting smooth curves alternates between smoothing by running medians and eye fitting smooth curves to points. The particular running median smoothing recommended is to first take medians of 5 and then to take medians of 3 of these (medians of 5) where this makes a difference.

To fit a simple curve, we follow the sequence:

- smooth by medians.
- plot results, eye fit, find residuals from eye fit.
- plot results, eye fit, find residuals from eye fit.
- add up eye fits, plot, smooth again (if needed), adjusting residuals accordingly.

In the example of section 29yD, this process is repeated four times (the last step being omitted in each of the first two) in order to make a fit of the form

$$\text{function}_1(u - v) + \text{function}_2(u + v)$$

where $\text{function}_1(\ )$ and $\text{function}_2(\ )$ are smooth -- and graphically given when we are done -- and where the heights to be fitted are given for the combination of certain u's with certain v's. (In the example u and v are column cut and row cut flogs going with the cuts fixing the double flog being fitted.)

We use four full cycles in order to fit functions of each of u - v and u + v twice.

We also illustrate, in section 29yE, the proper use of a tilt on log smallest corner when this is to be done after a column-plus-row fit. The essential is to use, as a carrier, the residual of log smallest corner after it, too, has been fitted by columns-plus-rows.

****

29yP. Problems for chapter 29y

1 to 32. Do the like-numbered problems of exhibit 8.

33. Do problem 1 of exhibit 12.

34. Do problem 2 of exhibit 12.

35. Do the problem of exhibit 16.

36. Do the problem of exhibit 19.

37. Do problem 1 of exhibit 22.

38. Do problem 2 of exhibit 22.

39. to 42. Do the project of exhibit 22.

43. Do problem 1 of exhibit 23.

44. Do problem 2 of exhibit 23.

45. Do problem 1 of exhibit 24.

46. Do problem 2 of exhibit 24.

47. Do problem 3 of exhibit 24.